WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Detection of Malware in Android Smartphones Using Machine Learning

Jayanti Kamewari Surya Naga Bharati

B. Tech Student, Department of IT, GMR Institute of Technology, Rajam-532127, Andhra Pradesh, India
Email: 21341A1253@gmrit.edu.in

ABSTRACT

Now-days smartphones are becoming very famous all over the world. As the study says, among all platforms, Android is the widely used platform. The widespread adoption of Android smartphones has drawn the attention of malware designers, which threatens the Android ecosystem. This research presents an approach that utilizes Machine Learning (ML) techniques to detect viruses on Android devices. Machine Learning models, such as Support Vector Machines (SVM) Random Forests are trained using preprocessed data. These models learn the patterns and characteristics of both virus infected and healthy devices. To evaluate the effectiveness of this approach extensive experimentation and cross validation techniques are employed. The models undergo testing under scenarios, including known virus samples, zero day attacks and false positives generated by applications. The ML-based methods for detecting source code vulnerabilities are discussed, because it might be more difficult to add security after the app is deployed. This paper help researchers to have a general idea of the malware detection approaches, pros and cons of each detection approach, and methods that are used in these approaches.

Keywords: *Machine Learning (ML) and Deep Learning (DL), Android malware, code vunerabilities, Mobile security*

INTRODUCTION

Android smartphones have become an integral part of our daily lives, offering convenience and connectivity. However, with their widespread use, they have also become prime targets for malicious actors who aim to compromise user data, privacy, and device security through Android malware. These threats come in various forms, such as viruses, Trojans, spyware, and ransomware, posing a significant risk to users' digital well-being. Machine Learning (ML) has emerged as a formidable ally in the battle against Android malware. ML leverages the power of artificial intelligence to analyze extensive datasets, recognize evolving patterns, and make informed decisions. This technology empowers security experts to not only detect known malware but also predict, identify, and respond to emerging and adaptive threats in real-time. The dynamic nature of Android malware requires an equally adaptable defense, making ML an invaluable tool.This introduction highlights the pressing need for robust security measures in the Android ecosystem and sets the stage for a deeper exploration of the symbiotic relationship between Android malware and Machine Learning, underscoring how ML is reshaping the landscape of smartphone security.

RESEARCH APPROACH

The paper have been conducted in the field of Android malware detection, focusing on different aspects such as accuracy, dataset size, feature sets, and machine learning algorithms.

The paper gives the study of ensemble learning techniques such as XGBoost and AdaBoost to detect the malicious apps,achieving high accuracy.

Deep learning has been shown to be successful in classifying Android malware, especially with additional training data .

SVM-based malware detection techniques have also been explored, incorporating dangerous permission combinations and susceptible API calls as elements in the SVM algorithm .

Permission-based strategies and feature selection using the Genetic algorithm have also been used for Android malware detection .

Some studies have addressed the challenge of model sustainability and the multicollinearity problem in machine learning models.

METHODOLOGY:

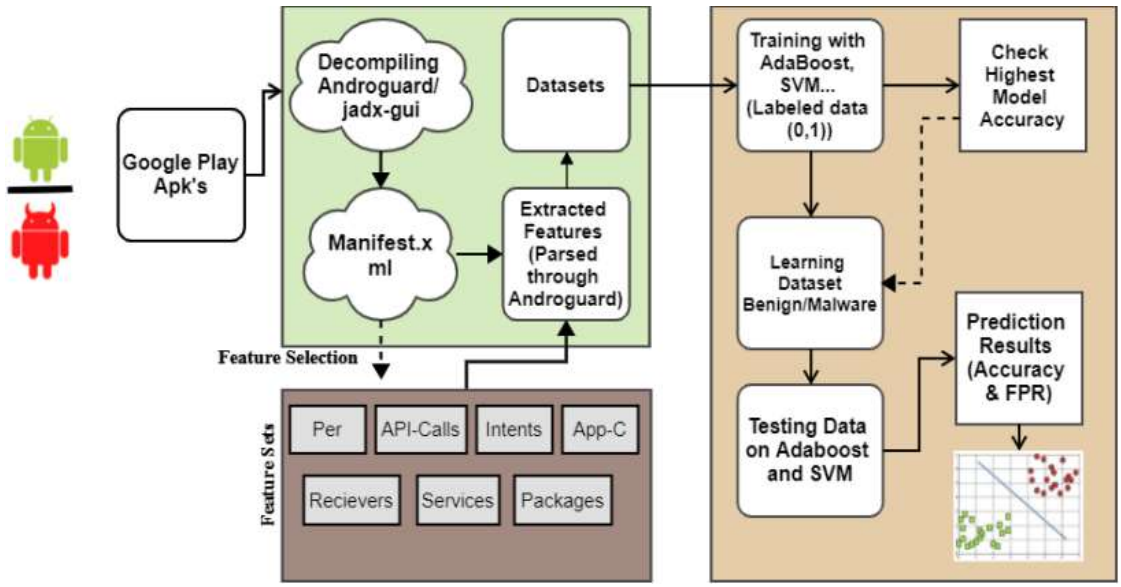

Fig : Proposed methodology of our system.

We have taken up the task to train up to six machine learning algorithms such as AdaBoost, Support Vector Machine, Decision Tree, KNN, Navies Bayes and Random Forest techniquesand classify these machine learning algorithms accurately. The methodology section is divided in two sections; Pre-Processing (explaining the prerequisite processing) and the Proposed Model (Model functionalities and components).

Data Preprocessing: APK files from numerous apps were included in the resulting datasets (containing malware and benign characteristics). A Jadx-Gui decompiler is then used to reverse engineer the apk files to extract features from the Android manifest file's feature set for further processing. These stages are regarded as pre-processes from before real assessments and are essential parts before any kind of testing and training using any predictive models.

Train and Test of data: Three different datasets were used for training and testing in this paper, including apps from Google Play, MalDroid, DefenseDroid, and a generated dataset .The datasets were combined to incorporate multiple feature sets for higher accuracy and classification of malware .The datasets were first trained on every algorithm for comparative classification analysis .The accuracy of the algorithms was evaluated, and the dataset was again trained and tested on the higher-performing algorithms .The trained models were used to predict the output for a given Android application's extracted features .The results showed an accuracy of 96.24% in detecting extracted malware from Android applications, with a 0.3 False Positive Rate (FPR).

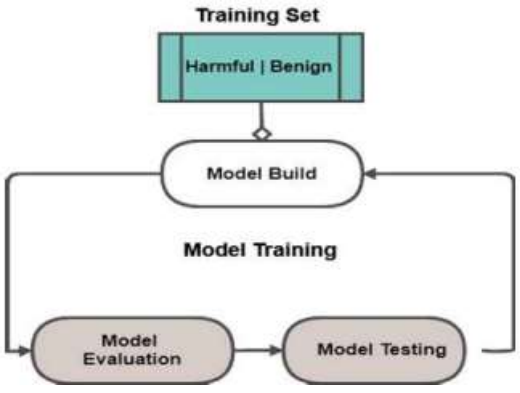

Fig :Training and Testing of dataset

The training cycle of the program and how the model first is constructed and then evaluated. Then further on the data is cycled towards testing and that is the data fed to the trained model for further prediction analysis of the android applications.

Random Forest Classifier: Random Forest Classifier was one of the machine learning algorithms used in this research paper for malware detection in Android applications . It was selected as one of the six models to experiment with, along with other algorithms such as AdaBoost, Naive Bayes, Decision Tree classifier, K-Neighbor, and Gaussian NB.The Random Forest Classifier was trained and tested on the datasets containing features and malware samples to classify Android applications as either benign or malware.

Results

Model	Features	Datasets	Malware	Accuracy	FPR
AdaBoost	55821 (Selected)	MalD+DefenseD+GD	18578	96.24%	0.301%
AdaBoost	55821 50621+331 56741 (Selected)	MalD+GD	12931	95.74%	0.416%
SVM	55821 331+ 56471 (Selected)	MalD+ DefenseD +GD	18578	92.04%	0.731%
SVM	331 (Selected)	GD	5877	90.1%	0.970%

CONCLUSION

In conclusion, the integration of Machine Learning (ML) into malware detection has revolutionized the field of cybersecurity. The adaptability and learning capabilities of ML algorithms make them highly effective in identifying and thwarting a wide range of malicious software. The improved accuracy of ML-based detection systems, reducing false positives and negatives, enhances the overall reliability of cybersecurity measures. The automation facilitated by ML streamlines threat detection processes, enabling real-time analysis and response. The emphasis on behavioral analysis allows ML models to detect novel threats and zero-day attacks by identifying abnormal patterns in system behavior. The scalability of ML-driven solutions makes them suitable for protecting networks of varying sizes, while the reduced time to detection is critical for minimizing the impact of cyber threats. Furthermore, the continuous improvement inherent in ML, through updates based on evolving threat intelligence, ensures that these systems remain robust against emerging challenges. However, it is essential to recognize that ML is most effective when integrated as part of a comprehensive cybersecurity strategy, complementing other security measures and staying abreast of the ever-changing cyber threat landscape. As research and development in ML techniques continue, the synergy of advanced technologies will play a pivotal role in fortifying our defenses against sophisticated and evolving cyber threats

References

[1] Senanayake, Janaka, Harsha Kalutarage, and Mhd Omar Al-Kadri. "Android mobile malware detection using machine learning: A systematic review." Electronics 10, no. 13 (2021): 1606.

[2] Alzaylaee, M., & Alzaylaee, M. (2021). Android Malware Detection using Machine learning: A Review. International Journal of Computer Science and Information Security, 19(1), 7-13.

[3] Xu Jiang, Baolei Mao, Jun Guan, Xingli Huang, 2020," Android Malware Detection Using FineGrained Features", Scientific Programming, vol. 2020, Article ID 5190138.

[4] A. O. Christiana, B. A. Gyunka, and A. Noah, ''Android malware detection through machine learning techniques: A review,'' Int. J. Online Biomed.Eng., vol. 16, no. 2, p. 14, Feb. 2020, doi: 10.3991/ijoe.v16i02.11549.

[5] Statista_a (2019) Number of available applications in the Google Play Store from December 2009 to December 2019.

[6] Senanayake, J., Kalutarage, H., & Al-Kadri, M. O. (2021). Android mobile malware detection using machine learning: A systematic review. Electronics, 10(13), 1606.

[7] Mahindru, Arvind, and A. L. Sangal. "MLDroid—framework for Android malware detection using machine learning techniques." Neural Computing and Applications 33, no. 10 (2021): 5183-5240.

[8] Kakavand, Mohsen, Mohammad Dabbagh, and Ali Dehghantanha. "Application of machine learning algorithms for android malware detection." Proceedings of the 2018 International Conference on Computational Intelligence and Intelligent Systems. 2018.

[9] Niveditha, V. R., and T. V. Ananthan. "Detection of Malware attacks in smart phones using Machine Learning." International Journal of Innovative Technology and Exploring Engineering 9, no. 1 (2019): 4396-4400.

[10] Alzaylaee, M. K., Yerima, S. Y., & Sezer, S. (2017, March). Emulator vs real phone: Android malware detection using machine learning. In Proceedings of the 3rd ACM on International Workshop on Security and Privacy Analytics (pp. 65-72).